Practice FINAL

Dr. Nancy Pfenning
Statistics

The Final will be a closed book exam worth 250 points. You are allowed to bring a calculator and 2 two-sided sheets of notes.

1. (10 pts.) Select the most appropriate display for each of the following.
   (a) Workers' income and whether they are black, white, Asian, or Hispanic:
   (i) bar graph (ii) histogram (iii) pie chart (iv) scatterplot
   (v) back-to-back stemplot (vi) side-by-side boxplots
   (b) Rent charged and distance from the university for a sample of Oakland apartments:
   (i) bar graph (ii) histogram (iii) pie chart (iv) scatterplot
   (v) back-to-back stemplot (vi) side-by-side boxplots
   (c) Workers' marital status and whether they smoke or not:
   (i) histogram (ii) bar graph (iii) scatterplot (iv) back-to-back stemplot
   (v) side-by-side boxplots

2. (5 pts.) Which one of the following is a paired study?
   (a) Measure level of depression for a sample of women who have had cosmetic surgery and for a sample of women who haven't had cosmetic surgery
   (b) Measure level of depression for a sample of women prior to undergoing cosmetic surgery, and again a year following the surgery

3. (5 pts.) In general, which is more likely to contain the unknown population mean?
   (a) a 90% confidence interval (b) a 99% confidence interval (c) both the same

4. (25 pts.) Scientists believe people's ears get larger with age. They measured ear length in a sample of patients, aged 30 to 93, and found their ears grew about .01 inches a year.
   (a) Was this an experiment or an observational study?
   (b) What is the explanatory variable? Is it quantitative or categorical?
   (c) What is the response variable? Is it quantitative or categorical?
   (d) According to our notation, .01 is denoted by which of the following?
   $\mu, \sigma, p, \beta_0, \beta_1, \chi^2, n, \bar{x}, s, \hat{p}, b_o, b_1, \alpha$
   (e) A Chinese physician recalled his mother's childhood nagging: "Stretch your ears daily, child, to ensure long life." In fact, some scientists believe that people with small ears die younger, leaving a population of healthier old people with big ears. What would be the explanatory variable, according to this theory?

5. (30 pts.) Select the most appropriate statistical test for each of the following; choose from

   (i) z test about a proportion (ii) z test about a mean with one-sided alternative (iii) z test about a mean with two-sided alternative (iv) t test about a mean with one-sided alternative (v) t test about a mean with two-sided alternative (vi) two-sample t test with one-sided alternative (vii) two-sample t test with two-sided alternative (viii) chi square test (ix) ANOVA (x) inference for regression

   (a) We survey a random sample of Oakland households to test if "nuclear families" (father, mother, offspring) are in a minority there:
   (i) (ii) (iii) (iv) (v) (vi) (vii) (viii) (ix) (x)

   (b) We look at the mean and standard deviation of a small random sample of Oakland one-bedroom apartments to decide if the overall mean rent could be 500 dollars:
   (i) (ii) (iii) (iv) (v) (vi) (vii) (viii) (ix) (x)

   (c) We want to test if the mean rent of all Oakland one-bedroom apartments is less than the mean rent of all Shadyside one-bedroom apartments:
   (i) (ii) (iii) (iv) (v) (vi) (vii) (viii) (ix) (x)

   (d) We test the IQ of a large random sample of children of women smokers to see if their mean IQ is significantly lower than the national mean of 100. We assume population standard deviation to be 16:
   (i) (ii) (iii) (iv) (v) (vi) (vii) (viii) (ix) (x)

   (e) We examine a random sample of Oakland apartments to see if overall there is a relationship between rent charged and size (in square feet):
   (i) (ii) (iii) (iv) (v) (vi) (vii) (viii) (ix) (x)

   (f) We take random samples of married, single, and divorced workers to determine if mean earnings differ among these groups:
   (i) (ii) (iii) (iv) (v) (vi) (vii) (viii) (ix) (x)

   (g) We want to test for a relationship between race and employment status (employed or not):
   (i) (ii) (iii) (iv) (v) (vi) (vii) (viii) (ix) (x)

6. (25 pts.) Below are a scatterplot and regression output, where the data consist of average viewer ratings (on a scale of 1 to 10) and gross receipts (in millions of dollars) of 15 recent movies:

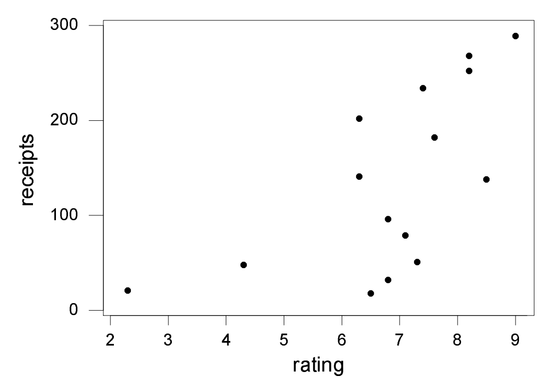
(a) Considering the appearance of the scatterplot, which of the following is the most reasonable guess for the correlation r?
(i) -.95 (ii) -.6 (iii) -.1 (iv) +.1 (v) +.6 (vi) +.95

(b) Use the regression equation to predict the receipts of a movie rated 9 by viewers. ____________UNDERLINE the part of the output that estimates about how far off this prediction would be.

(c) According to the regression output, is there statistical evidence of a relationship between ratings and receipts? Answer yes or no and CIRCLE THE SPECIFIC PART OF THE OUTPUT THAT YOU USE TO DECIDE.

(d) Since estimated receipts for a movie rated 8 are $-113 + 36.5(8) = 179$, and $s = 75.93$, a 95% confidence interval for mean receipts of movies rated 8 is roughly $179 \pm 2(75.93)/\sqrt{15} =$ **(140, 218)** and a 95% prediction interval for receipts of a particular movie rated 8 is roughly $179 \pm 2(75.93) =$ **(27,331)**. Are you willing to believe that all movies rated 8 average 300 million in receipts?

(e) If we set up a prediction interval for receipts of particular movie rated 4 instead of 8, what would be the most noticeable difference: (i) the interval would be centered at lower values (ii) the interval would be wider (iii) the interval would be narrower




```
Regression Analysis: receipts versus rating
The regression equation is
receipts = - 113 + 36.5 rating
```

Predictor	Coef	SE Coef	T	P
Constant	-112.60	84.37	-1.33	0.205
rating	36.45	12.00	3.04	0.010

S = 75.93 R-Sq = 41.5% R-Sq(adj) = 37.0%

7. (30 pts.) This table gives results consistent with those of a study that compared the incidence of prematurely gray hair for people in their early thirties who were or were not addicted to drugs or alcohol.

	Gray	Not Gray	Total
Addicted	106	94	200
Not Addicted	130	670	800
Total	236	764	1000

(a) The data were probably obtained from (i) an observational study (ii) an experiment (iii) it could easily have been either (i) or (ii).

(b) State the appropriate null and alternative hypotheses.

(c) Which group has a higher proportion with gray hair?

(d) Complete the table of counts expected under the null hypothesis.

Expected	Gray	Not Gray	Total
Addicted			200
Not Addicted			800
Total	236	764	1000

(e) Calculate the value of the chi-square statistic. ____________

(f) How many degrees of freedom are there?

(g) The P-value is (i) very small (ii) somewhat small (iii) borderline (iv) not small at all

(h) Draw your conclusions (circle **two** of the following):

i. There is evidence of a relationship between addiction and gray hair.

ii. There is no compelling evidence of a relationship between addiction and gray hair.

iii. These people could represent random samples taken from populations where equal proportions had gray hair for addicts and non-addicts.

iv. We do not believe these workers represent random samples taken from populations where equal proportions had gray hair for addicts and non-addicts.

8. (5 pts.) A biologist observes a sample of 81 cells to have mean dividing time 30.5 minutes and standard deviation 4.9 minutes.

(a) Construct a 95% confidence interval for population mean dividing time μ.

(b) Suppose a sample of only 11 cells had mean 30.5 and standard deviation 4.9. Give three reasons why a 99% confidence interval for μ would be wider than your interval in (a).

9. (20 pts.) For several years, a mathematics placement test has been administered to all incoming freshmen at a certain college. Those in charge of administering the test claim the average time required is no more than 50 minutes, but students suspect the average time is in fact more than 50 minutes. A random sample of 4 students selected over this time period took on the average $\bar{x} = 54$ minutes to complete the test, with a standard deviation $s = 3$ minutes.

   (a) Test at the $\alpha = .05$ level, making sure to identify the appropriate hypotheses, test statistic, range for the p-value based on the graph below, p-value size (small, not small, or borderline), and conclusion: is there compelling evidence that the mean time exceeds 50 minutes? (yes or no or not sure)

   (b) Give a range for the p-value if a test were made against a **two-sided** alternative.

   (c) Suppose the P-value for the one-sided test is calculated to be 0.04. Which one of the following is correct?

      i. The probability that the test requires on the average 50 minutes is no more than 4%.

      ii. The probability that the test takes longer than 50 minutes is 4%.

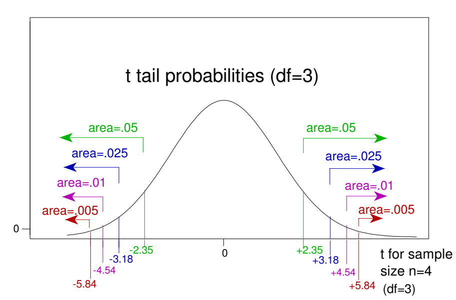
      iii. If the average time required to take the test is 50 minutes, then the probability of the sample producing an average time **at least as long as** 54 minutes is 4%.

      iv. If the average time required to take the test is 50 minutes, then the probability of the sample producing an average time of 54 minutes is **at least** 4%.

   (d) If there is statistical evidence that the test requires more than 50 minutes, then the administration must go through a great deal of trouble to revise the test. Students, on the other hand, could benefit from having them create a shorter test.

      i. Who would be most interested in avoiding a Type I error, administrators or students?

      ii. Who would be most interested in avoiding a Type II error, administrators or students?



10. (20 pts.) Researchers at the University of California Earthquake Center want to compare intensities of earthquakes at San Francisco Bay and at the Los Angeles Basin. The data, assumed to be independent random samples from the two areas, represent intensity on the Richter scale of tremors felt during several months in 1994.

    (a) If researchers want to test if mean intensities differ between the two regions, state the appropriate null and alternative hypotheses.

    (b) Use the two-sample t statistic in the output to characterize the p-value: (i) extremely small (ii) somewhat small (iii) borderline (iv) not small at all

    (c) Is there statistically significant evidence of a difference?

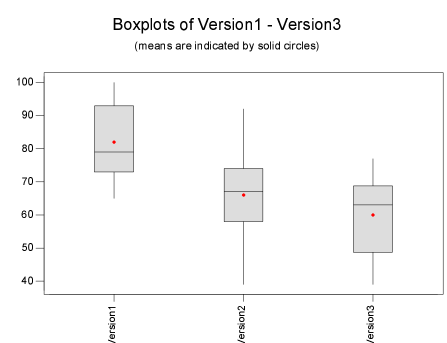
    (d) Would you expect a confidence interval for difference in population means to contain zero?

    (e) Are your conclusions still valid if one of the data sets is very skewed?

    (f) Circle the part of the output that would help you decide if the Rule of Thumb for use of a pooled procedure is satisfied.

```
Two-sample T for SF vs LA
                N       Mean        StDev      SE Mean
SF             14       2.59        1.02         0.27
LA             16       2.13        1.09         0.27
Difference = mu SF - mu LA
Estimate for difference: 0.454
%95% CI for difference: (-0.339, 1.248)
T-Test of difference = 0 (vs not =): T-Value = 1.18
```

11. (10 pts.) Below are boxplots for students' scores on 3 different versions of an exam. Tell whether each of the following would make the data more convincing or less convincing that in general the three exams do not share the same level of difficulty. (a) if sample means differed more: (i) more convincing (ii) less convincing (b) if the boxes were taller, with longer whiskers (i) more convincing (ii) less convincing (c) if sample sizes were larger (i) more co



12. (30 pts.) Three groups of six guinea pigs each were injected, respectively, with small, medium, and large doses of a new tranquilizer, and the following are data for the number of minutes X_i (for $i = 1, 2, 3$) it took them to fall asleep:

$$\begin{array}{rlll} \text{small:} & n_1 = 6 & \bar{x}_1 = 22.5, & s_1 = 2.168 \\ \text{medium:} & n_2 = 6 & \bar{x}_2 = 20.0, & s_2 = 1.414 \\ \text{large:} & n_3 = 6 & \bar{x}_3 = 13.0, & s_3 = 2.098 \end{array}$$

(a) Is this study an experiment? (i) Yes, because more than one dosage level is tested. (ii) Yes, because there is a treatment involved (injecting with various amounts of tranquilizer) for the units. (iii) No, this is an observational study on a random sample of guinea pigs. (iv) No, because there is no control group of guinea pigs which receive no injection.

(b) Under which dosage level did sample guinea pigs fall asleep fastest?

(c) Is it safe to assume that population standard deviations are equal (as required in an analysis of variance)? Answer yes or no, and circle the values above that you use to decide.

(d) State the appropriate null and alternative hypotheses.

(e) Complete the blanks in the ANOVA table below.

(f) Circle **two** of the following which are correct conclusions: (i) Mean time to fall asleep may be the same for all guinea pigs given the three dosage levels. (ii) Not all the mean times to fall asleep are the same for populations of guinea pigs on the three dosage levels. (iii) There is a relationship between dosage level and time it takes to fall asleep. (iv) There is no relationship between dosage level and time it takes to fall asleep.

(g) Which of these would tend to lead to **less** convincing evidence? (You may circle anywhere from none to all three.)

  i. if sample sizes were larger
  ii. if sample standard deviations were larger
  iii. if sample means were more different

Source	DF	Sum of Squares	Mean Square	F	P-value
Group		291.0			0.000
Error		55.5		XXXXXXX	XXXXXXX
Total	XX	XXXXX	XXXXXXX	XXXXXXX	XXXXXXX

(h) Which of these is the best way to assign guinea pigs to treatments?

  i. with eyes closed select 6 guinea pigs to receive small doses, then 6 guinea pigs to receive medium doses; the rest receive large doses
  ii. use random numbers from a table or generated by software
  iii. give small doses to the 6 least active guinea pigs, and large doses to the 6 most active
  iv. give small doses to the 6 most active guinea pigs, and large doses to the 6 least active

13. (20 pts.) In a test for Extra Sensory Perception, a subject chooses each time from 4 images the one that is supposed to be "sent" to him telepathically. Out of 60 questions, he gets 18 correct.

   (a) If he is really only guessing, then the **sample count** X of correct answers is a binomial random variable with $n = 60$ and $p = .25$. Its mean is ______________and standard deviation is ______________.

   (b) The z-score for 18 correct is ______________; the probability of scoring at least 18 is (i) not small at all (ii) borderline (iii) somewhat small (iv) extremely small

   (c) To formally test if the subject has ESP, we can test if his **proportion** of correct answers is significantly higher than it would be if he were only guessing. Under the null hypothesis of random guessing, the overall proportion of correct answers is $p = .25$. For a sample of 60 questions, sample proportion $\hat{p}$ would have mean ______________and standard deviation ______________.

   (d) The p-value is the probability of observing a sample proportion $\hat{p}$ at least as high as 18 out of 60. The p-value is
   (i) not small at all (ii) borderline (iii) somewhat small (iv) extremely small

   (e) Is there statistical evidence that the subject has ESP?

   (f) Construct a 90% confidence interval for his long-run proportion of correct answers, based on the sample proportion correct; does the interval contain .25?

14. (15 pts.) Recently, warning labels covering 30% of each cigarette pack were made mandatory in all European Union countries. Although the U.S. pioneered tobacco warning labels in 1965, it has not upgraded its cautions since 1984; they cover less than 20% of a pack, and are fairly unobtrusive. In Canada, warnings on cigarette labels cover 50% of the front and back of each pack, and include full-color pictures of organs ravaged by smoking. Researchers would like to know the impact of the various labels on people's smoking habits.

   (a) What would be the most obvious pitfall in a study which would compare percentages smoking in Europe, the U.S., and Canada in order to determine the impact of the 3 types of label? (i) confounding variables (ii) placebo effect (iii) volunteer bias (iv) response bias (v) non-response

   (b) Suppose smokers in Europe, the U.S., and Canada were asked to rate the impact of the labels on their desire to smoke, from 0 = "no impact" to 10 = "they convinced me to quit". Would this be an experiment or an observational study? To test for a significant difference among impacts depending on whether a smoker was exposed to the labels in Europe, the U.S., or Canada, researchers should use (i) z test about a proportion (ii) paired t (iii) two-sample t (iv) ANOVA (v) regression (vi) chi-squared

   (c) In a random sample of 400 Canadian smokers, 44% said the new warnings increased their desire to quit. Construct a 95% confidence interval based on these results. Which one of the following is a correct interpretation of your interval?

      i. 95% is the probability that the confidence interval produced from this sample contains the proportion of all Canadian smokers whose desire to quit would be increased by the warnings.
      ii. 95% is the probability that the proportion of all Canadian smokers whose desire to quit would be increased by the warnings falls in the interval.
      iii. 95% is the probability that the confidence interval produced from this sample contains the proportion of all smokers whose desire to quit would be increased by the warnings.
      iv. 95% is the probability that the proportion of all smokers whose desire to quit would be increased by the warnings falls in the interval.

   (d) Based on your confidence interval in (c), what would be the outcome of a test of $H_o : p = .5$ vs $H_a : p < .5$ where p refers to the population proportion whose desire to quit would be increased by the warnings?